DOCUMENT RESUME

ED 041 933 TM 000 019

AUTHOR Kurfman, Dana G.
TITLE Using Questionnaire Data to Revise Curriculum Materials.
PUB DATE Mar 70
NOTE 6p.; From symposium "Evaluation of New Social Studies Curriculum Projects" (Annual meeting American Educational Research Association, Minneapolis, Minn., March 1970)

EDRS PRICE EDRS Price MF-$0.25 HC-$0.40
DESCRIPTORS *Curriculum Development, *Curriculum Evaluation, Curriculum Research, Evaluation Criteria, Evaluation Techniques, Geography Instruction, Instructional Materials, Questionnaires, *Social Studies, Symposia
IDENTIFIERS High School Geography Project

ABSTRACT

The need to evaluate course materials developed by the High School Geography Project led to the development of an evaluation strategy. Test data was of little use in identifying problems or in suggesting revisions in curriculum, but questionnaire information was useful in rating the effectiveness of the material. There are four critical stages in this evaluation strategy: the organizational stage, (identification of the key elements); the criteria stage (questionnaires on interest, effectiveness, significance, clarity, and sufficiency); interpretation (when judgments must be made about levels of positive response); and, finally, directive reporting of the results that makes clear the changes needed in the materials. Although the effectiveness of this strategy can be assessed by obtaining positiveness scores on a revised unit, a more significant possibility would be the development of norms by which to interpret questionnaire responses. However, the first method has indicated gains in student interest on 14 activities, all but 3 being significant. (PR)

1970 AERA SYMPOSIUM: Evaluation of New Social Studies Curriculum Projects

Using Questionnaire Data to Revise Curriculum Materials

by

Dana G. Kurfman

The High School Geography Project has been developing geography course materials since 1964. Throughout this period the Project invested significant resources in the evaluation of the materials. The results of these evaluation efforts were specific recommendations for revising or re-forming the materials.

School tryouts on a national scale were held for thirteen geography units during four school years. Several units were tried out once, revised, and tried out again. Usually about twenty-five teachers and 800 students were involved in each tryout. All teachers were paid volunteers and some were included in more than one tryout. They were undoubtedly better prepared in their subject than most high school geography teachers. For most tryouts the median verbal aptitude of students was about the 60th percentile. About two-thirds of the students were in ninth or tenth grade--the grade levels for which the materials were designed.

Test and questionnaire data were obtained during each tryout of materials. Test data proved to be of little use in identifying problems in the materials or in suggesting revisions. There are two reasons for this. One is the fact that objectives were usually stated explicitly only after the educational activities had been developed. The second reason is the difficulty of interpreting data derived from pretest and posttest administrations of multiple choice items.*

Questionnaire data provided the basis for most of the recommendations transmitted by HSGP evaluators to the revision editors. Questionnaire data are normally used to provide testimonials and creative suggestions for improving the materials. This paper describes procedures of a different nature in which questionnaire information was used primarily to rate the effectiveness of the materials in an objective manner. Each part of the materials was judged in terms of such criteria as interest, clarity, and educational worth, as perceived by teachers and students. These ratings provided a major basis for deciding whether to retain, revise, or discard each educational activity or unit.

This paper reports four critical stages in an evaluation strategy using questionnaire data to revise course materials. These can be called the organizational stage, the criteria stage, the interpretation stage, and the reporting stage. It then suggests a procedure for assessing the effectiveness of the evaluation strategy itself.

*An elaboration of these problems will appear in a forthcoming paper.

Before turning to this formative evaluation strategy, it is necessary to say something about the HSGP materials that were evaluated. HSGP materials consist of student readings and other resources, guides for the teacher indicating what to do and how to do it, and assorted media to be used in the learning activities. Thus, materials imply both media and procedures for using the media. Teacher's Guides are both directive and detailed. The materials are divided into units requiring from four to eight weeks of teaching time. Each unit consists of parts called activities. Generally, there are five to ten such activities in a unit.

The first critical stage in formative evaluation is identification of the key parts of the materials. This is important because these parts become the focus of student and teacher opinions. In the HSGP evaluation strategy, careful attention was given to distinguishing and naming the activities making up each unit. In the first years of development and revision, Project evaluators learned that it was possible to divide a unit into too many activities, many of which had no unique characteristics. Consequently, activities often were indistinguishable to students several days later. If students and teachers fail to have the intended activity in mind when stating their opinions, obviously the essential validity of these opinions is seriously in doubt. Thus, one of the first requirements of effective formative evaluation is the organization and labelling of activities, readings, and films so that students will remember them.

The criteria stage is a major feature of every evaluation model. Evaluative criteria are implicit in the questions asked about each activity, film and reading. Five criteria were used during each HSGP evaluation: interest, effectiveness, significance, clarity, and sufficiency.

The interest criterion applies to students. They were asked simply to indicate their degree of interest in each activity compared to typical school experiences. Although a number of scales would be suitable, a four point scale from "dull" to "very interesting" was used. Student interest questions were also used for the readings and other special features of each activity.

The effectiveness criterion was used by asking teachers to judge the effectiveness of each activity in terms of the teaching materials and procedures they normally use. This criterion was also applied to more specific features of the materials, such as the effectiveness of questions in stimulating discussion or the effectiveness of various kinds of visuals as educational tools. Simple "yes-no" responses to such questions as, "Is the film effective in stimulating discussion?" were sought.

The significance criterion was used by asking both students and teachers to indicate the importance or worth of what was learned in each activity. They were asked to make this judgment on a four point scale in terms of other school learnings.

As might be expected, the clarity criterion was applied to a number of aspects of the materials. These include the clarity of directions for each activity, the clarity of the reading for different levels of student capability, the clarity of the objectives as stated for an activity, and the clarity of test questions used. The sufficiency criterion was applied to the amount of reading asked of students, the amount of time suggested for each activity and the extent of the geographic background provided for teachers in helping them to teach an activity. "Yes-no" responses were sought for clarity and sufficiency questions.

It is clear that such criteria as interest, effectiveness, significance, clarity, and sufficiency are needed as a basis for questions that provide the data for formative evaluation. Responses to questions on a four point scale can then be turned into means, or percentages in the case of "yes-no" responses. These means and percentages indicate degrees of positive response to the questions.

The interpretive stage in an evaluation strategy requires judgments about levels of positive response. Activities that do not meet acceptably positive levels need to be revised. The liklihood that volunteer teachers may be positive about almost anything is well established. Determining acceptable levels of response to any question is a difficult judgment to make. In the beginning it is probably best simply to compare activities in terms of the response to each question. Thus, if 70 to 90 per cent of the teachers in a tryout respond positively to most activities, but only 30 per cent respond positively to one activity, this suggests the advisability of revising that activity. As experience develops, degrees of positiveness can be interpreted to be low, average, and high. Over several years of HSGP data collection 80-95 per cent positive came to be considered a high degree of student interest, 65-80 per cent an average degree, and less than 65 per cent a low degree of student interest. Although it varied in terms of the criterion a high degree of teacher positiveness was over 90 per cent, an average degree was 75-90 per cent, and a low degree was something under 75 per cent. In this manner rough norms for interpreting questionnaire data can be developed.

Judgments about the interest level, effectiveness, significance or clarity of an activity are worthwhile to the extent that they are made by students and teachers representative of the ultimate users of the material. When such judgments are made by extraordinary teachers and students, the results may be suspect. The opinions of typical students and teachers are most useful in identifying weaknesses in the materials, even though they may provide little help in suggesting how to correct the weaknesses.

At the reporting stage of formative evaluations, it is not sufficient simply to report the means and percentages obtained. Their meaning for revising the materials must be made as clear and directive as possible. Otherwise, the quantity of data can be so great that the people charged with revising the materials are often unable to interpret them. What is needed is a very careful interpretation on the part of evaluators, always indicating the criteria used in reaching a conclusion. It has

been HSGP experience that the results of formative evaluation should be reported in a way that makes as explicit as possible the changes that seem to be needed in the materials.

The evaluation strategy used by HSGP to try out and revise course materials thus has four characteristics. One is a careful identification of the segments to be evaluated; a second is the formulation of questions based on five criteria; a third is the interpretation of degrees of positive response; and a fourth is the directive reporting of results.

There is no generally accepted method of assessing the effectiveness of such a strategy. However, one way of doing so is to try out the revised materials again. When a revised unit is tried out again, positiveness scores can be obtained on the same measures that were used originally. Effective revision presumably would lead to a more positive result on the second tryout.

There are problems inherent in attempting to compare results from one tryout to another. Ideally, random samples would be drawn from the teacher population for each tryout. There is no reason that this could not be done by large school districts interested in implementing a systematic evaluation of curriculum products for use in its schools. Since volunteer teachers were used in HSGP tryouts, however, no claim is made for their representativeness. In point of fact, correlational analyses on more than one occasion indicated that no teacher characteristic significantly affected the results obtained from either student or teacher questionnaires.

This "test" of a set of evaluation procedures could be made in terms of a number of criteria. Only student interest is reported here because many of the other questions were changed during the four year period. Moreover, student interest correlates highly with other criteria that could be used.

The following table indicates the degree of positiveness of student interest from "dull" to "extremely interesting" for an earlier and a later tryout of a number of activities. Each activity was revised in terms of an evaluation report based on the earlier tryout. A response of "dull" is assigned a one and a response of "extremely interesting" is assigned a four. Thus, any mean rating over 2.50 means that more than half of the students rated the activity positively in terms of interest. Data of this sort are available only for the activities indicated. Other activities were simply discarded after the first tryout, or combined with other activities so that it was not clear that essentially the same activity was being evaluated on both occasions. The table shows only those activities which remained essentially the same from the first to the second tryout. The number of students varied from unit to unit and tryout to tryout.

Student Estimates of Interest in Selected HSGP Activities

	1966-67 or 1967-68	1968-69	Increase
Geography of Cities Unit			
Site Diagrams	2.83	3.13	.30**
Portsville	3.34	3.65	.31**
Models of City Form	2.72	2.75	.03
Manufacturing & Agriculture Unit			
Game of Farming	3.59	3.65	.06
Hunger	2.79	3.13	.34**
Interviews with Farmers	2.85	3.00	.15*
Agricultural Realm	2.73	2.83	.10*
Metfab	3.09	3.26	.17**
Geographic Patterns of Manufacturing	2.78	2.84	.06
Geography of Culture Change Unit			
Cattle	2.85	2.97	.12**
Games	2.85	3.00	.15**
Sports	2.81	3.16	.35**
Culture Change	2.69	3.02	.33**
Canada	2.51	2.85	.34**

* significant at the .05 level
** significant at the .01 level

For the activities on which data are available HSGP's formative evaluation program seems to have led to notable increases in student interest. All activities show some increase. The mean for all fourteen activities at the end of the first tryout was 2.88. At the end of the second tryout it was 3.09. Only three are not significant at an acceptable (.05) level. Six are significant at the .01 level. Thus, improved questionnaire results can be used to assess the effectiveness of questionnaire based evaluation just as improved test performance can be used to assess the effectiveness of test based evaluation.

If representative samples of teachers are used, it should be possible to determine the effectiveness of an evaluation strategy in the manner indicated. Of more practical significance is the possibility of developing norms by which to interpret the questionnaire responses of teachers and students. School districts would then have a way of evaluating the effectiveness and interest level of curriculum materials like educational games, films, readings and resource units. Decisions to accept, reject or modify such materials could then be made in terms of objective data and not just the intuitive judgments of a few teachers and administrators.